Affinity CNN: Learning Pixel-Centric Pairwise Relations for Figure/Ground Embedding

Michael Maire
TTI Chicago

Takuya Narihira
Sony Corp.

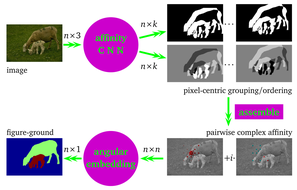
Stella X. Yu
UC Berkeley / ICSI

Overview

Goal: Segmentation and figure/ground from a single image, solved jointly

Approach:

- CNN predicts pairwise relationships
- New complex-valued affinity model:
  segmentation (grouping), figure/ground (ordering) $\} \rightarrow W$
- Angular Embedding for globalization
- Performance leap over prior work [1]



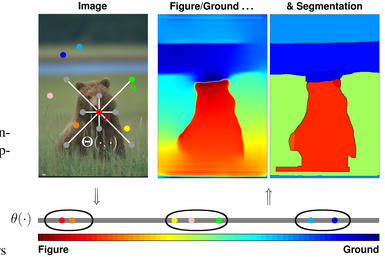
Segmentation and Figure/Ground via Embedding

Embedding:

- $\Theta(p,q)$ is an offset
  - Is p in the same region as q?
  - Is p in front of/behind q?
- $C(p,q)$: confidence on $\Theta(p,q)$
- Recover $\theta(\cdot)$: $\Theta(p,q) \approx \theta(q) - \theta(p)$
- $\theta(\cdot)$ is a globally consistent representation of the pairwise local grouping/ordering relationships $\Theta(\cdot,\cdot)$

Implementation:

- Learn to *directly* estimate C, Θ
- Sparse multiscale stencil: each pixel p connects to k neighbors



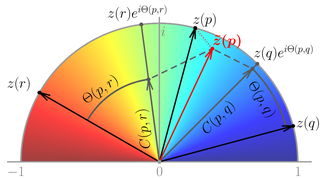
Angular Embedding [2]

Given:

- Pairwise ordering relationships $\Theta(\cdot,\cdot)$
- Confidence on each relationship $C(\cdot,\cdot)$

Recover:

- Global ordering $\theta(p)$ by embedding:
  $p \rightarrow z(p) = e^{i\theta(p)}$



Minimize:

$$\varepsilon = \sum_p \frac{\sum_q C(p,q)}{\sum_{p,q} C(p,q)} \cdot |z(p) - \tilde{z}(p)|^2$$

where:

$$\tilde{z}(p) = \sum_q \bar{C}(p,q) \cdot e^{i\Theta(p,q)} \cdot z(q)$$

$$\bar{C}(p,q) = \frac{C(p,q)}{\sum_q C(p,q)}$$

Relax to generalized eigenproblem:

$$Wz = \lambda Dz$$

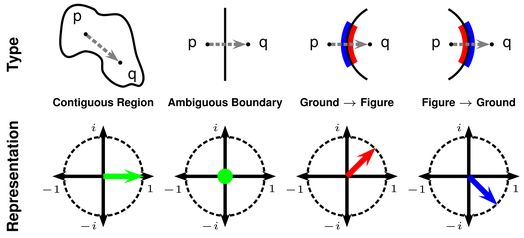
where:

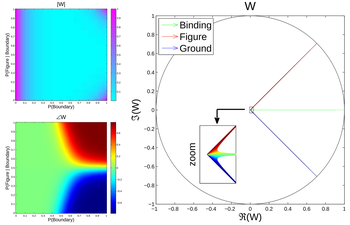
$$D = \mathrm{Diag}(C1_n) \qquad W = C \bullet e^{i\Theta}$$

(Normalized Cuts is the special case $\Theta = 0$)

[1] M. Maire. Simultaneous Segmentation and Figure/Ground Organization using Angular Embedding. *ECCV*, 2010.
[2] S. X. Yu. Angular Embedding: A Robust Quadratic Criterion. *PAMI*, 2012.

Complex Affinity Model





Basic Relationship Types:

- Contiguous region
  - p and q lie in the same region
  - high confidence ($C(p,q) = 1$) on zero displacement ($\Theta(p,q) = 0$)
- Ambiguous boundary
  - p and q lie in different regions
  - no local cues for relative displacement $\hookrightarrow$ low confidence ($C(p,q) = 0$)
- Figure transition
  - positive displacement ($\Theta(p,q) = \phi$)
- Ground transition
  - negative displacement ($\Theta(p,q) = -\phi$)

Pairwise Pixel Interactions:

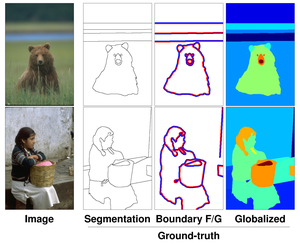
- Probability estimates:
  $e(p) = Pr(p \text{ lies on a boundary})$
  $b(p,q) = Pr(\mathrm{seg}(p) \neq \mathrm{seg}(q))$
  $f(p,q) = Pr(\mathrm{figural}(p,q) \mid \mathrm{seg}(p) \neq \mathrm{seg}(q)))$
  $g(p,q) = Pr(\mathrm{figural}(q,p) \mid \mathrm{seg}(p) \neq \mathrm{seg}(q)))$
- Transition error probabilities:
  $E_B(p,q) = b(p,q)$
  $E_F(p,q) = 1 - (1 - e(p))b(p,q)(1 - e(q))f(p,q)$
  $E_G(p,q) = 1 - (1 - e(p))b(p,q)(1 - e(q))g(p,q)$

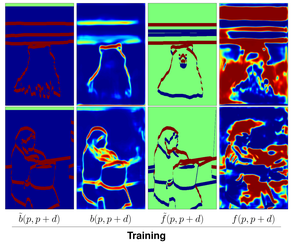
B: binding **F**: ground → figure **G**: figure → ground

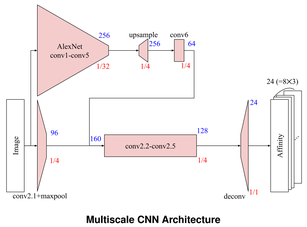
Affinity from Interactions:

- Convert probabilities to confidence:
  $C_B(p,q) = \exp(-E_B(p,q)/\sigma_b)$
  $C_F(p,q) = \exp(-E_F(p,q)/\sigma_f)$
  $C_G(p,q) = \exp(-E_G(p,q)/\sigma_g)$
- Apply rotational actions:
  $W_B(p,q) = C_B(p,q)$
  $W_F(p,q) = C_F(p,q)\exp(i\phi)$
  $W_G(p,q) = C_G(p,q)\exp(-i\phi)$
- Sum transition forces and symmetrize:
  $W = W_B + W_F + W_G$
  $W \leftarrow (W + W^*)/2$

Affinity Learning from Globalized Ground-truth







Annotated Ground-truth:

- Human-drawn segmentation (boundary map)
- Figure/ground along region boundaries only

Globalized Ground-truth:

- Desire: learn to predict pairwise figure/ground between possibly spatially distant pixels p and q
- Solution: apply Angular Embedding to the ground-truth, recovering a globalized version!

Training Input:

- Segmentation map $S(\cdot)$: pixel → region id (from annotated ground-truth)
- Figure/ground rank $R(\cdot)$: pixel → $\mathbb{R}$ (from globalized ground-truth)

Training Targets:

$$\tilde{b}(p,q) = 1 - \delta(S(p) - S(q))$$

$$\tilde{f}(p,q) = (sign(R(q) - R(p)) + 1)/2$$

Learning:

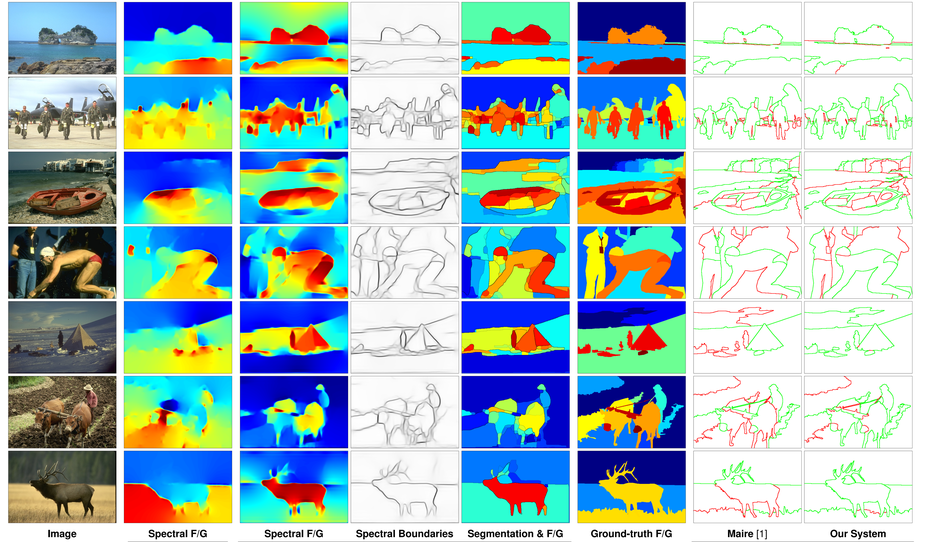
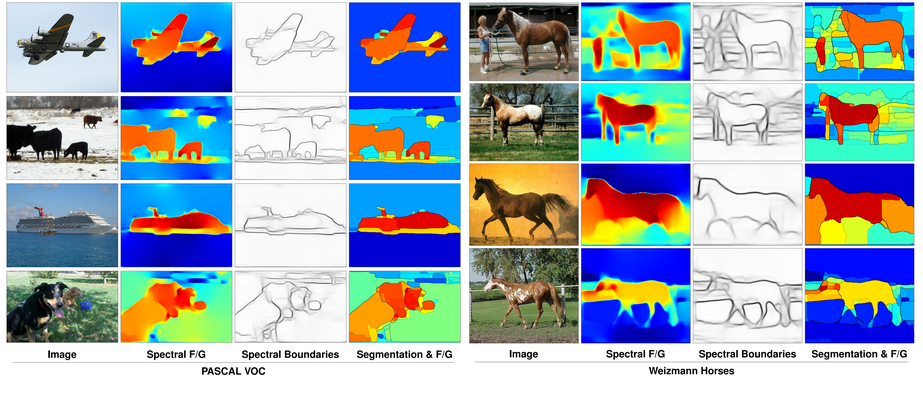
- CNN output: $b(\cdot,\cdot)$, $f(\cdot,\cdot)$ maps per neighbor (stencil with 8 offsets × 3 scales = 24 neighbors)
- SGD with log loss averaged pixel-wise:

$$\mathcal{L}^{\log}(\tilde{y},y) = -\tilde{y}\log(y) - (1-\tilde{y})\log(1-y)$$

$$\mathcal{L}_b = \frac{1}{|I|}\sum_{p\in I}\frac{1}{|\mathcal{N}(p)|}\sum_{q\in\mathcal{N}(p)}\mathcal{L}^{\log}(\tilde{b}(p,q), b(p,q))$$

- Data augmentation: translate, left-right mirror

Results and Evaluation on Berkeley Segmentation Dataset



Cross-Domain Generalization